

# IARIW-ESCoE Conference

## “Measuring Intangible Assets and Their Contribution to Growth”

---

### A First Look at Open Source Software Investment in the United States and in Other Countries, 2009-2019

*Carol A. Robbins, National Center for Science and Engineering Statistics, National Science Foundation, crobbins@nsf.gov*

*Gizem Korkmaz, Biocomplexity Institute, University of Virginia, gkorkmaz@virginia.edu*

*Ledia Guci, National Center for Science and Engineering Statistics, National Science Foundation, lguci@nsf.gov*

*José B. Santiago Calderón, Biocomplexity Institute, University of Virginia, jbs3hp@virginia.edu*

*Brandon L. Kramer, Biocomplexity Institute, University of Virginia, brandonleekramer@gmail.com*

Disclaimer: The views expressed in this paper are those of the authors and do not necessarily reflect the position of the respective institutions.

#### ABSTRACT

Researchers and policymakers can access detailed data about patents, designs, and research publications as output indicators of useful knowledge. However, as digitization of the tools of knowledge increases, this picture is increasingly incomplete. Digitization also provides the opportunity to gather data about other types of useful knowledge and knowledge assets [1].

In this paper, we develop new measures of intangible capital. Specifically, we use GitHub repository data on software freely shared between 2009 and 2019 to develop time series estimates of annual nominal and real investment and real capital stocks for open source software (OSS) in the United States. We estimate investment in OSS by using a cost approach to produce measures comparable to other nonmarket intangible investment, such as capitalized R&D, entertainment, literary, and artistic originals, and own-account business software. Our estimates for OSS on GitHub in 2019 place the equivalent cost of OSS created by U.S. contributors (contributors with U.S. addresses) at over \$300 billion with an equivalent cost of global OSS creation on GitHub over \$500 billion (based on U.S. costs). The richness of the collected GitHub data enables further analyses, including the measurement of contributions in OSS by various economic sectors within the U.S. and countries and the study of structural features of international collaborations within the global OSS ecosystem.

Open Source Software (OSS), as defined by Open Source Initiative, is computer software with its source code shared with a license in which the copyright holder provides the rights to study, change, and distribute the software to anyone and for any purpose. OSS is developed, maintained, and extended both within and outside of the private sector, through the contribution of independent developers as well as people from universities, government research institutions, businesses, and nonprofits. Many OSS projects are hosted in free repositories, and information on contributors and development activity embedded in these repositories is publicly available. GitHub is the largest

platform with over 50 million users and developers worldwide. We collect 7.8 million project repositories, containing metadata such as author, license, commits (approved code edits), and lines of code.

The methodology for estimating investment through resource costs for OSS development follows standard methodologies for accounting for nonmarket output based on the sum of all costs. We use lines of code as the input measure of effort to estimate the time spent on software development, accounting for project complexity. Wages and compensation are based on Bureau of Labor Statistics occupation-level data. Non-labor costs are estimated based on industry account ratios from the U.S. input-output tables. The capital stocks are created with the perpetual inventory method using computer software depreciation rates from the Bureau of Economic Analysis. The estimates presented in this paper extend previous work on the resource cost of creating open source software packages for the R and Python software languages [2, 3].

Finally, the GitHub data, which we plan to make publicly available, presents a unique opportunity to conduct supplementary analyses around the development of open source software. Two important applications are presented in this paper. First, using the same resource cost approach, we examine the contribution of the government sector to OSS and present estimates of the amount of OSS shared by the U.S. federal government between 2009 and 2019. Second, we use OSS contributors' locations to generate contributor networks [4] and study structural features of international collaborations using social network analysis methods. We also identify key players in the OSS ecosystem using various network centrality metrics.

## References

- [1] Keller, S., G Korkmaz, C. Robbins, S Shipp. Opportunities to observe and measure intangible inputs to innovation: Definitions, operationalization, and examples. *Proceedings of the National Academy of Sciences* 115 (50), 12638-12645, 2019.
- [2] Robbins, C., G. Korkmaz, J. Calderon, C. Kelling, S. Shipp, and S. Keller. The scope and impact of open source software: A framework for analysis and preliminary cost estimates. In *The 35th IARIW General Conference: The Digital Economy-Conceptual and Measurement Issues*. The International Association for Research in Income and Wealth (IARIW), 2018. <http://www.iariw.org/copenhagen/robbins.pdf>.
- [3] Robbins, C., G. Korkmaz, J. Calderón, D. Chen, A. Schroeder C. Kelling, S. Shipp, S. Keller. The Scope and Impact of Open Source Software as Intangible Capital: A Framework for Measurement with an Application Based on the use of R Packages. National Bureau of Economic Research Conference on Research on Income and Wealth, Bethesda MD, March 15-16, 2019.
- [4] Korkmaz, G.; C. Kelling; C. Robbins; & S. Keller. (2019). "Modeling the Impact of Python and R Packages Using Dependency and Contributor Networks." *Social Network Analysis and Mining (SNAM)*. 10:7. <https://doi.org/10.1007/s13278-019-0619-1>